# LINGUISTIC PITCH ANALYSIS USING FUNCTIONAL PRINCIPAL COMPONENT MIXED EFFECT MODELS

JOHN A. D. ASTON[1,2], JENG-MIN CHIOU[2], AND JONATHAN E. EVANS[3]

Abstract. Fundamental frequency (F0, broadly "pitch") is an integral part of human language; however, a comprehensive quantitative model for F0 can be a challenge to formulate due to the large number of effects and interactions between effects that lie behind the human voice's production of F0, and the very nature of the data being a contour rather than a point. This paper presents a semi-parametric functional response model for F0 by incorporating linear mixed effects models through the functional principal component scores. This model is applied to the problem of modelling F0 in the tone language Qiang, a language in which relative pitch information is part of each word's dictionary entry.

---

*Date*: October 13, 2008.

*Key words and phrases.* Phonetics; Fundamental Frequency; Random Effect Models; Principal Component Analysis; Functional Response Models.

[1]CRiSM, Dept of Statistics, University of Warwick, UK

[2]Institute of Statistical Science, Academia Sinica, Taiwan

[3]Institute of Linguistics, Academia Sinica, Taiwan

Address for Correspondence:

John Aston

Centre for Research in Statistical Methodology

University of Warwick

Coventry CV4 7AL

UK

Tel: +44-(0)-24-7657-4808

Fax: +44-(0)-24-7652-4532

Email: j.a.d.aston@warwick.ac.uk.

## 1. Introduction

Phonetics is the branch of Linguistics relating to the study of the sounds produced during speech. Each spoken language has particular sound patterns and properties which are inherent to that language, and which form a system that is somewhat independent from the grammatical organisation of words and their meaningful components. These features include sound segments such as consonants and vowels, as well as suprasegmental properties of duration, pitch, and intensity for example. The aim of this paper is to adapt and apply current statistical semi-parametric curve estimation methods for functional data to the analysis of linguistic pitch. This will allow investigation into the properties of speech sounds to a much more complex and quantitative degree than has previously been considered. Due to there being both fixed and random covariates associated with the model, the analysis will be achieved through the combination of linear mixed effects (LME) models and functional principal component analysis (FPCA).

Many quantities are of interest when investigating speech, such as duration of segments, intensity, and vowel quality. However, of particular interest in many studies is the fundamental frequency (F0, roughly "pitch"). From the articulatory (physiological) perspective, F0 is the number of complete cycles of vibration of the vocal cords measured in Hertz (Crystal 1990). From an acoustic (sound) perspective, a speech signal is a complex periodic wave composed of multiple sine waves. The frequency of repetition of this complex wave is its F0 (Johnson 1999, p. 10)

At the syllable level, F0 can be modelled either as a point or as a curve. Models which are based on a single point per syllable either use a summary statistic (Khouw and Ciocca 2007; Evans, Chu, and Aston 2008) or a target value (Beckman and Hirschberg 1994). Models that are based on the F0 excursion over the syllable take within-speaker averages (Rose 1987; Xu 1999; Stanford 2008) in order to have smoother, more "typical" curves to compare. Curves are typically time-normalized, and often smoothed, before averaging, as in Xu (1999).

Other curve-based models depend on predefined contour models (Fujisaki and Hirose 1984). Acoustic studies of F0 tend to either rely on invariant syllable structure (Xu 1999; Xu 2006; Fujisaki, Gu, and Hirose 2004), or ignore the measurements at the edges of the vowel, in order to reduce the effects of syllable-initial and syllable-final consonants (Mixdorff 2000). Studies often trim as much as 10% of the beginning and end of the vowel; in more unusual cases, as much as 25% of the beginning may be trimmed (Stanford 2008).

While these methods of analysis can make the models easier to consider, there are major drawbacks in that speakers produce and listeners perceive the entire contour, and thus have it available to them while interpreting the sounds they are hearing. In addition, models that are based on a single type of syllable cannot be extended to other syllable types, and those that intentionally remove the effect of consonants are not able to predict complete F0 trajectories. Thus, from the perspective of both production and perception, these models are limited in their applications. In some languages, such as tonal languages, relative pitch contours may be part of each word's dictionary entry and thus be necessary for both fluent pronunciation and for comprehension. Therefore for the model to be interpreted as a more appropriate model for pitch, the output should consist of contours as opposed to point estimates. Some studies have included analysis of the speech contour (Xu 1999; Xu and Xu 2003), but have required extensive assumptions relating to the data, such as invariant syllable structure, and often the reading of nonsense words to have a complete experimental design for the purpose of averaging. However, in many spoken languages, including the example in this paper, no written form exists: speakers cannot read, and will refuse to utter, nonsense words. In addition, speech patterns vary from person to person, and as such, a model needs to take into account this random nature.

In order to combine all these effects together, a simple semi-parametric functional response model will be proposed. A FPCA will be performed on the pitch contours to extract component curves which are present in the data. The resulting associated functional principle component (FPC) scores, which determine how much of each principal component curve is present in each observation, will then be used as the response variable in a parametric LME model, to account for all the covariates both of a fixed and random nature that might be present in the data. This modelling approach has the advantages of not requiring prespecification of the pitch contours present. This is especially important as it cannot be known a-priori exactly what contour shapes will be present, yet it is of interest to try to associate particular contours with particular covariates. The use of FPCA with LME allows a large number of covariates to be included in the model for the way the contours are combined. The overall aim of this paper is to propose a method to find a linguistic description of the pitch information in language through both the curve and coefficient estimates.

The rest of the paper is organised as follows. A brief introduction to pitch analysis is given in the next section. Section 3 introduces the model and outlines how the combination of FPC scores and LME models will be used for its estimation. Section 4 contains a small simulation study on the finite sample properties of the FPC estimation in a similar context to the experimental data. Section 5 outlines the application of the model introduced to a tonal dialect of the Qiang language of Sichuan Province in Mainland China. The final section gives some concluding remarks and discussion of the relevance and possible extensions of the model. The appendices expand upon some details of the combination of FPCA and mixed effect models.

## 2. Pitch Analysis

In languages with stress (e.g., English), pitch, or equivalently F0, is often an integral component of stress marking, as in *ˈe-le-ˌva-tor ˈo-pe-ˌra-tor*, in which the pitches of the syllables can follow a relative height pattern of 4-1-2-1 3-1-2-1 (Trager and Bloch 1941). In English and many other languages, stress is also indicated by other factors such as intensity, syllable duration, and vowel quality changes. This combination can be observed in the phonetic differences between *REcord* (noun) and *reCORD* (verb). Due to the number of effects that indicate stress, the pitch pattern of stress can be altered for effect, so that in *Did* *you say "elevator operator"?*, the first syllable of "elevator" may be lowered, yet still convey stress.

In a neutral utterance of the aforementioned compound, "operator" starts at a slightly lower F0 than "elevator", although both initial syllables carry primary stress. Across the world's languages, phrases and statements generally start at a higher pitch than they end on, with a relatively smooth slope downward from start to finish; questions may have a dramatic pitch rise at the end, etc. Phrase-level pitch patterns like these are termed intonation. Thus, a stressed syllable at the end of a sentence may occur at a lower F0 than an unstressed syllable at the beginning of the same sentence. From this fact it can be seen that pitches in language are produced and perceived relative to those of nearby syllables, and are not defined by exact frequency, unlike pitch in music, where the note A above middle C has been standardised at 440 Hz (ISO 16).

Half or more of the world's languages have at least some morphemes (words or meaningful sub-parts of words) in which pitch specification is an integral component; this component is called "tone". Using a relative scale ranging from low to high, tone contrasts in Mandarin Chinese may be represented as follows:

| | | | |
|---|---|---|---|
| ma˥ | "mother" | ma˧˥ | "tingle" |
| ma˨˩˦ | "horse" | ma˥˩ | "scold" |

where the tone marks represent approximate contours for changes in pitch. Changing the pitch pattern on a syllable changes the vocabulary item that is being said. Like stress, tone is subject to intonation, so that a high tone that is later in an utterance may have lower F0 than an earlier low tone.

Aside from tone, stress, and intonation, numerous linguistic and non-linguistic properties can influence F0. These include sex of the speaker, type of sentence, preceding/following tones/stress, properties of preceding/following consonants, and the vowel being said. In addition, the speaker him-/herself is a random effect: his/her customary pitch range, size of vocal cords, health condition, etc., all contribute to F0. Not only are these effects important contributors to F0, they also may interact in significant ways.

In addition, language communities combine the universally available effects in unique ways; e.g., Japanese women speak at higher pitches than do Dutch women (Van Bezooijen 1995). The linguist is challenged to model the way that speakers of a given language combine the effects at their disposal to produce F0 in a manner that is consistent with their speech community.

For many more remote speech communities it is difficult to get large quantities of data, and thus the model must be able to make the best use of all available data. It is also unrealistic to expect people to speak nonsense words or phrases in order to achieve a balanced design covering all possible sound interactions so as to be able to average out their effect; since such words and phrases are inherently unusual, they can cause speakers to alter their speech patterns in unusual ways.. Thus any reasonable model for F0 should be able to include many covariates and interactions, be based solely on natural speech, and also allow for the fact that the data is really a contour over time.

## 3. Statistical Methodology

The analysis of contours and curves is now well established in the statistics literature; for many examples see Ramsey and Silverman (2002, 2005) and Ferraty and Vieu (2006). In particular, since Castro, Lawton, and Sylvestre (1986) and Rice and Silverman (1991), the nonparametric estimation of the mean and covariance function has developed into the area of FPCA. The incorporation of random effects into functional data has also received some attention in the literature. Several basis function methods have proposed to account for the mixed effects including those based on either smoothing spline approaches (Guo 2002) or wavelet based approaches (Morris and Carroll 2006). For the phonetic analysis considered here, it is important to minimise assumptions on the shape of the curves, and the use of nonparametric curve estimation helps achieve this objective. Methods have recently been developed for hierarchical FPCA random effects models (Di, Crainiceanu, Caffo, and Punjabi 2008), but due to the large number of covariates that likely affect the data, including emphasis on the modelling of random subject effects, neither the hierarchical nor the single-index modelling approach as in Chiou, Müller, and Wang (2003) can be easily considered. Instead, a mixed effect parametric model for the FPC scores and the covariates is considered. This has the intrinsic advantages of being able to account for and test easily the influence of the covariates, and also allows the relatively easy interpretation of the results back in the domain of interest to the phonetician, despite the non-parametric specification of the curves themselves.

Let $Y_i(t)$, $t \in T = [0, 1]$, $i = 1, \ldots, n$ be data sampled from a Gaussian stochastic process on the domain $T$. While $T$ often represents time, in this study, $T$ represents vowel time, from the beginning to the end of the vowel. This normalisation (time warping) of vowels into a synchronised time frame is often used in phonetic analysis, as it allows curves to be considered across a common time scale, even though different instances of vowels last different lengths of time. For each sample process $Y_i$, two sets of scalar covariates $X_i$ and $Z_i$ are available. $X_i$ are fixed effects, such as tone, while $Z_i$ are zero-mean Gaussian random effects, such as speaker. The following model is proposed:

$$\begin{aligned} E(Y_i(t)|X_i, Z_i) &= \mu(t) + \sum_{j=1}^{\infty} E(A_{i,j}|X_i, Z_i)\phi_j(t), \\ E(A_{i,j}|X_i, Z_i) &= X_i\beta^{(j)} + Z_i\gamma^{(j)}, \quad \gamma^{(j)} \sim N(0, \Sigma_{\gamma^{(j)}}), \end{aligned} \tag{1}$$

where $\phi_j(t)$ is the $j$th basis function and $A_{i,j}$ is the weight associated with the $i$th curve and the $j$th basis function. $\mu(t)$ is the overall mean of the sampled processes. Essentially the process is modelled as a mean function coupled with a stochastic basis expansion component. The $A_{i,j}$ are modelled as LME with fixed effect coefficients $\beta^{(j)}$ and random coefficients $\gamma^{(j)}$.

The analysis to find the FPC eigenfunctions $\phi_j(t)$ follows the methodology developed by Chiou, Müller, and Wang (2003). In fact, the basis functions in (1) were chosen to be the eigenfunctions in the data which can be estimated from the empirical covariance matrix.

While all the elements in the decomposition can be smoothed as required, as this was not required in the example, this has been omitted as the data already looked quite smooth. Let $t_{i,j}$, $j = 1, \ldots, m_i$ be the time points for the $i$th subject. In the example, the sampling is the same for all $i$, and thus the $i$ index of $t_{i,j} = t_j$ and $m_i = m$ will be omitted henceforth.

An estimate of the mean function $\hat{\mu}(t_j)$ can be simply calculated from the mean of the data. The eigenfunctions are then determined from a spectral analysis of the estimated covariance matrix

$$\hat{C}(t_k, t_l) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i(t_j) - \hat{\mu}(t_j)\right) \left(Y_i(t_l) - \hat{\mu}(t_l)\right), \quad k, l = 1, \ldots, m \tag{2}$$

This yields the estimated eigenfunctions $\hat{\phi}_j(t)$ as

$$\hat{C}(t_k, t_l) = \sum_{j=1}^{m} \lambda_j \hat{\phi}_j(t_k) \hat{\phi}_j(t_l) \tag{3}$$

with ordered eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. The FPC scores $A_{i,j}$ are then estimated by discrete approximation

$$\hat{A}_{i,j} = \sum_{k=1}^{m} \left(Y_i(t_k) - \hat{\mu}(t_k)\right) \hat{\phi}_j(t_k) \Delta_k \tag{4}$$

where $\Delta_k = t_k - t_{k-1}$. In a similar way to traditional principal component analysis, each eigenfunction explains the maximum amount of variance of the stochastic process about its mean and all previous eigenfunctions, and thus the eigenvalues allow a measure of the proportion of explained variance. The estimation of the effect of the covariates on the $A_{i,j}$ is then carried out using a standard LME model analysis (Pinheiro and Bates 2000; Faraway 2006).

Having estimated the eigenfunctions and the FPC scores, model selection for both the regression model and the number of retained eigenfunctions is required. Firstly, given the presence of both fixed and random effects, a parametric bootstrap is used to select the relevant covariates of interest for the LME model, when the effects are close to the boundary of significance given by the asymptotic standard error estimates. For each $j$, the LME modelling proceeded by starting with the model containing all possible effects and interactions that were possible for the data (and estimable) and then removing covariates which were deemed to be insignificant at the 5% level corrected for multiple comparisons across eigenfunctions. While this top down (backwards elimination) approach does not guarantee the optimal model, it is a flexible and moderately robust approach given that combinatorial optimisation of the model covariates is not feasible.

To determine the number, $K$, of eigenfunctions to be retained, percentage of explained variance is commonly used as a choice, but in this data, with so many covariates, it is necessary to determine whether their influence comes through an eigenfunction with only a small related explained variance. Thus, the number of components needed for the model was determined through two procedures. If the percentage variance explained is too small to even account for the explained variance of a small fraction of one curve, then all components below this value were discounted. In addition, all the remaining eigenfunctions were checked for covariate effects and discounted as noise independent of the experimental setup, if they were not related to any covariate of interest. In principle, this is a slightly iterative procedure, as the number of curves to be accounted for in the multiple comparison is determined by $K$, but in practice one iteration is often all that is required.

In analysing the LME model, it was decided to use a mixture of Maximum Likelihood (ML) and REstricted ML (REML) methods. ML was used for model selection, as the parametric bootstrap was used for model comparison in cases where the mean and variance indicated that the covariate was close to the boundary of being included or not (see Faraway (2006) for a description of the use of the parametric bootstrap in LME models). Having selected the model, the REML parameter estimates were used as these are unbiased. For a much more in depth discussion of the choice between ML and REML, see Searle, Casella, and McCulloch (1992) among others. Confidence intervals for the parameters were generated using highest posterior density estimates from the REML estimates parameters as suggested in previous standard LME model analysis in Linguistics (Baayen, Davidson, and Bates 2008).

It is worth noting at this point that the assumption of a Gaussian process is required for the combination of FPCA and LME modelling. It is well known that for known eigenfunctions, the FPC scores are approximately Gaussian distributed (see Appendix A for more specific details). In addition, even though the number of time points in the example is relatively few (11 points), the number of curves is large (over 1000 curves), and as such it is reasonable to make the assumption that the eigenfunctions estimated are consistent (see section 4 for small simulation on this point). Given these, it is implied that the estimated scores will be approximately Gaussian distributed as well. Even though the FPC scores have the property that $E(A_{i,j}) = 0$, the conditional expectation $E(A_{i,j}|X_i, Z_i)$ helps describe the influence of the covariates on the FPC scores, and hence on the expectation of the functional response model (1).

In addition, given the Gaussian assumption, the $A_{i,j}$ are independent of one another across $j$. This means that the component scores from each separate eigenfunction can be modelled without reference to the other scores, allowing easy modelling and explanation. A particular contour may only be associated with a small subset of the covariates, which could indeed enhance interpretation (as will be seen in the example).

The overall specification has several advantages. Firstly, the $A_{i,j}$, $j = 1, \dots, K$ can be seen as a dimension reduction model for $Y_i(t)$, which allows a simple specification of the effect of the covariates, $X_i$ and $Z_i$, on the data. They are assumed only to affect the data through the weight of each basis function. While this makes the modelling simpler, it also makes interpretation much easier. For a linguist who is interested in the effect of a covariate, it amounts to the quantity of a particular contour that is added to the mean data signal when that covariate is present. It also allows for specification of confidence intervals on the covariate estimates, through such methods as Highest Posterior Density estimates. The relative ease of inference and model selection could be particularly useful in comparison to non-parametric regression settings.

An additional advantage of specifying the model in the form above is that it can then also handle very general forms of covariate. Typically, non-parametric regression analysis requires assumptions about the smoothness on the covariates. However, many of the covariates of interest in linguistic studies are binary, indicating the absence or presence of a linguistic effect, such as stress on the syllable, or discrete over a small finite set, such as the number of tones or vowels in the phonological inventory. By adding the parametric assumption, it becomes relatively straightforward to handle mixed effects models with such covariate structures.

## 4. Simulation: Assessing estimation consistency of eigenfunctions in finite sample data

In order to assess the assumption of negligible errors when estimating the mean function and eigenfunctions when there are large numbers of curves but which only have relatively few time points, the following simulation was undertaken. The simulation parameters were based on the linguistic data to correspond to the data analysis; in total, the data set consisted of 1386 F0 contours. The eight speaker's normalised F0 contours over the quadrisyllable /ɕí tʂ̩ú ˈpiàn tsə̀/ ("riverbank") is shown in Figure 1, as an example of the type of curve that was used to generate the simulation parameters.

1000 simulation samples of 1386 values were drawn from each of the LME models for the three FPC component scores, resulting in $\tilde{A}_{i,j}^{(m)}$, $i = 1, \ldots, 1386$, $j = 1, \ldots, 3$ and $m = 1, \ldots 1000$. As the scores are independent between eigenfunctions, these samples were drawn independently across $j$ and $m$. The sample scores were then centred due to the fact that the random effects can cause a slight shift in the mean away from zero and FPC scores have zero mean by construction. Simulated curves $\tilde{y}_i^{(m)}(t)$ were then generated from a linear combination of the mean function $\hat{\mu}(t)$, $\tilde{A}_{i,j}^{(m)}$ with $\hat{\phi}_j(t)$, $j = 1, \ldots, 3$ and noise proportional to the variance explained by the remaining eigenfunctions $\hat{\phi}_j(t)$, $j = 4, \ldots$. From the $\tilde{y}_i^{(m)}(t)$,

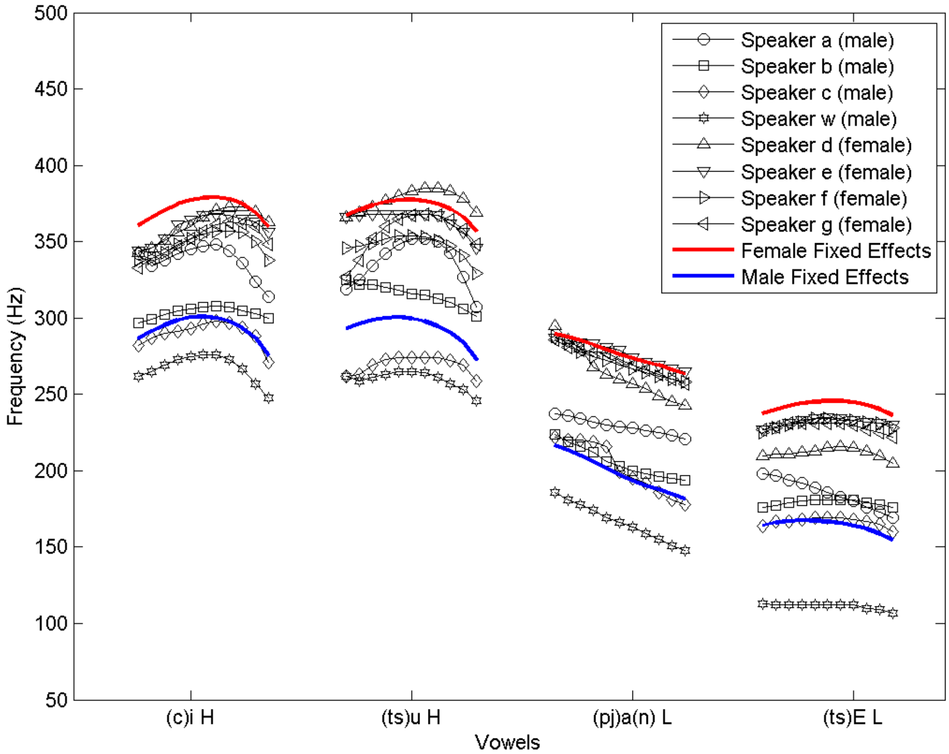



FIGURE 1. F0 contour curves for the quadrisyllable of /ɕí tʂú ˈpiàn tsə̀/ ("riverbank") for each of the four syllables for the eight speakers. The tonal pattern for the plotted data is *HHLL* and the sentence type is a declarative statement. The third syllable of the word is stressed. Also indicated are the estimated functional response model curves for males and females for the four syllables.

using the same procedure as described in Section 3, $\tilde{\phi}_j^{(m)}(t)$, $j = 1, \ldots, 3$ and $\tilde{\mu}^m(t)$ were estimated and compared with $\hat{\phi}_j(t)$, $j = 1, \ldots, 3$ and $\hat{\mu}(t)$, respectively.

Figure 2 contains the means of the estimated mean and eigenfunctions from the simulations along with empirical pointwise confidence interval estimates. As can be seen in the figure, there is very little variation in the estimates of either the mean function or the eigenfunctions from the simulations and these are overlapped by the estimated mean and eigenfunctions from the data. The greatest variation occurs at the end of the second eigenfunction where

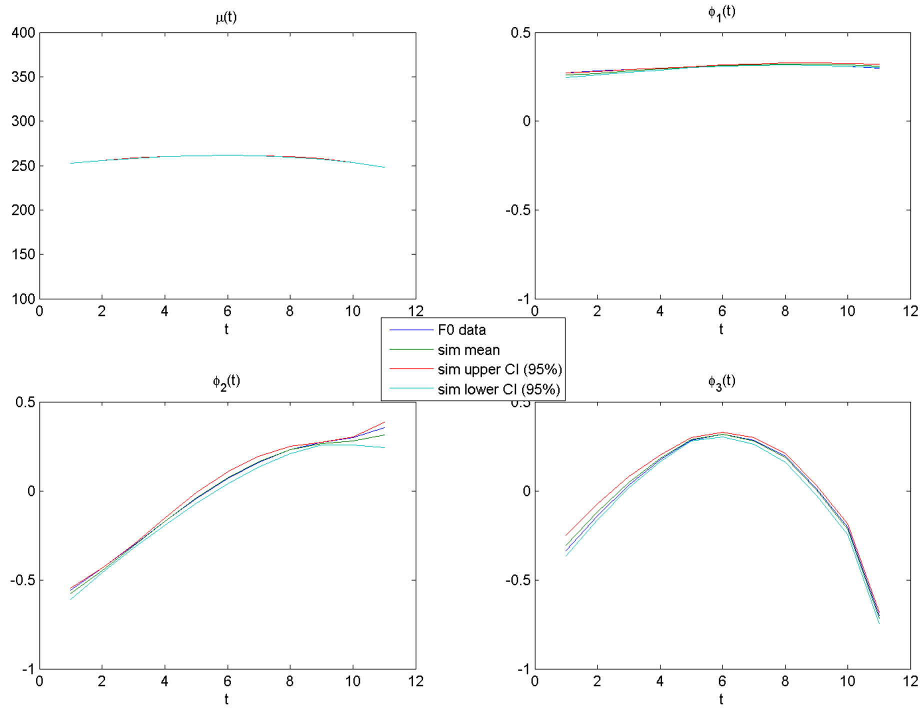



Figure 2. FPCA Simulations: Mean function and eigenfunctions for the first 3 components, along with the 95% confidence intervals. The simulation parameters were chosen to match the real data analysis. As can be seen, the true curves overlap the mean estimate and the confidence intervals are very small around the mean estimate.

the curvature is highest. However, even here, the variation is fairly limited. Overall, it would appear reasonable to make the assumption of negligible errors in the estimation of the mean and eigenfunctions for this data.

## 5. F0 Analysis of Luobuzhai Qiang

5.1. **Language Background.** The language studied is the Luobuzhai dialect of Qiang, a Tibeto-Burman language of Sichuan Province, China, with about 110,000 speakers (Liu 1998). The variety spoken in Luobuzhai village (about 1,000 speakers) is one of several Southern Qiang dialects, most of which demonstrate tone distinctions (Sun 1981; Evans

2001). The only published data on Luobuzhai come from Wen and Fu (1943); the data collected for this study appear to represent the first acoustic analysis of a Southern Qiang dialect.

Sun (1981) has asserted that the use of tone to distinguish lexical items is ubiquitous across Southern Qiang. However, the tone systems of Southern Qiang dialects are varied in their structure, and it is not always clear from published reports the role played by tone in each dialect. Constructing a quantified model of F0 would reveal the degree of importance of tone category in determining the fundamental frequency of syllables, and put that degree of importance in context with other factors that influence F0. The resulting model would provide a means of comparison with other Qiang dialects as well as other (tonal) languages, laying the groundwork for a quantified linguistic typology of F0.

A writing system for Northern Qiang has existed since 1993 (LaPolla and Huang 2003); however, due to the great differences in pronunciation and vocabulary between Northern and Southern Qiang, this writing system is not used in Southern Qiang dialects, such as Loubuzhai. Villagers who may be literate in Chinese are illiterate in Qiang. For this reason, some traditional elicitation methods, such as asking language consultants to read sentences or texts aloud into a microphone, are not available to the linguist studying this language. It is also not possible to have speakers of this language produce semantically anomalous expressions or nonsense words, which are used in many studies to fill out the data matrix. These methods can only be used among speech communities with a tradition of literacy.

5.2. **Data and Model Analysis.** The data set consisted of recordings of four male native speakers (ranging from 34 to 65 years old) and four female speakers (31 to 62 y.o.) gathered for an elicitation session in the home of one of the speakers. All of the speakers live in Luobuzhai village and use Luobuzhai Qiang as their most frequent mode of communication. The session took place prior to the 2008 Sichuan earthquake which devastated the region; about 200 residents of Luobuzhai died at that time, out of a population of around 1,000.

A list of nineteen nouns exemplifying the range of tonal and segmental variation was selected with the help of a native speaker. An attempt was made to find nouns whose tonal properties covered the widest possible range, and could fit within the same frame sentence, "I'm thinking about …". All example words were discussed in Chinese and in Qiang before being recorded. Because of an oral, rather than literate, culture, speakers had to find compounds acceptable before they would say them; semantic anomalies which fit the tonal patterns being sought were rejected by the subjects and were not recorded. The nouns were recorded within a frame sentence structure to yield three sentence types; statement, question or emphatic contrast. The list of the nouns used in the experiment is given in the appendix.

Pitch contours on vowels were identified via the software Praat (Boersma 1993). Syllable nuclei were sampled at eleven equidistant points, starting at the beginning of the vowel, at intervals of 10% duration, and at the end of the vowel. In this way, each syllable, regardless of duration, was sampled the same number of times (eleven).

Twelve possible variables (ten fixed, two random) were deemed to be of possible interest in the phonetic analysis of Luobuzhai Qiang. These include age, gender, tone, previous and following tones, sentence type (statement, question, or emphatic contrast), lexical stress (identified as the syllable containing the word's intensity peak), voicing of initial consonants, as well as the random effects of subject and word item. Not only were these effects considered separately, but up to third order interactions were also considered where linguistically appropriate. A full list of the covariates are given in Table 1.

The analysis was carried out in Matlab and R (R Development Core Team 2007). Matlab was used to find the eigenfunctions and FPC scores. The FPC scores were then modelled using the package lme4 (Bates and Sarkar 2007) for the mixed effects modelling in R, and the LanguageR package (Baayen 2007) was used to find the HPD confidence intervals using 50000 samples. Regression diagnostics were also performed using R.

| *Fixed Effects* | | |
|---|---|---|
| **Effect** | **Values** | **Meaning** |
| previous | #,H,L | Tone of previous syllable (# indicates word start) |
| tone | H,L | Tone of syllable |
| following | H,L | Tone of following syllable |
| condition | a,b,c | a=statement, b=question, c=emphatic contrast |
| gender | M,F | Gender of subject |
| vowel | a,e,i,u,ə | Vowel of Syllable |
| syll | linear | Position in word |
| voice | +,- | Initial consonant voiced |
| stress | +,- | Syllable stressed in word |
| age | linear | Age of subject |

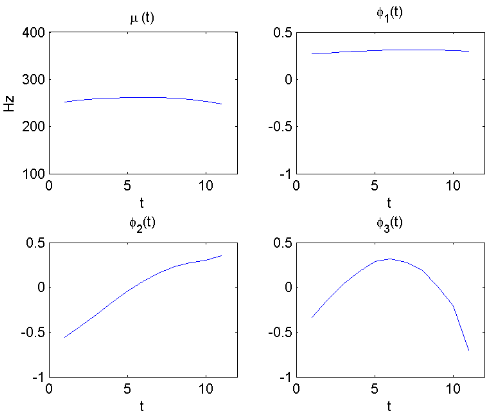
| *Random Effects* | | |
|---|---|---|
| **Effect** | **Value** | **Meaning** |
| subject | $N(0, \sigma^2_{\text{subject}})$ | Subject effect |
| word | $N(0, \sigma^2_{\text{word}})$ | Which word chosen effect |

TABLE 1. Covariates which have previously been linked with F0 production

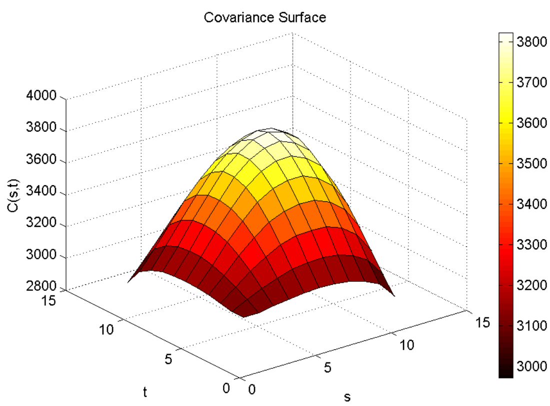
It was found that the Luobuzhai Qiang data was well modelled by taking $K = 3$ eigenfunctions. These were estimated from the empirical covariance matrix which was fairly smooth (see Figure 3), and thus additional smoothing was not deemed necessary. The first three eigenfunctions (see Figure 3) explained 99.8% of the variance of the data. In addition, all three models for the associated FPC scores contained meaningful covariates. The fixed effect covariate information for the models is given in Table 2 with the random effect covariates described in Table 3. It is of interest to note that the model for the first component explained

| Main Effect | FPC1 Estimate (l95,u95) | FPC2 Estimate (l95,u95) | FPC3 Estimate (l95,u95) | Interaction | FPC1 Estimate (l95,u95) | FPC2 Estimate (l95,u95) | FPC3 Esimate (l95,u95) |
|---|---|---|---|---|---|---|---|
| (Intercept) | 384.89 (315.32,455.12) | 16.26 (3.40,29.94) | -1.67 (-8.92,5.80) | previousH:toneL | -59.82 (-108.26,-11.06) | 3.22 (-11.87,18.14) | 24.51 (14.89,34.50) |
| previousH | 28.83 (-0.49,56.72) | -37.61 (-49.15,-26.28) | -13.80 (-20.06,-7.55) | previousL:toneL | -74.62 (-120.22,-32.82) | -19.97 (-34.00,-6.89) | 10.80 (1.76,19.93) |
| previousL | 35.06 (11.13,60.47) | 11.72 (1.82,22.25) | -8.70 (-14.23,-2.78) | previousH:followingL | -24.29 (-64.44,16.68) | -1.61 (-21.33,15.70) | 6.07 (-2.20,14.23) |
| toneL | -224.17 (-249.58,-197.93) | 1.37 (-7.43,9.80) | 2.08 (-4.23,7.88) | previousL:followingL | -11.11 (-49.70,27.02) | -28.60 (-48.22,-11.49) | 9.31 (1.21,17.54) |
| followingL | 95.89 (66.14,128.17) | 15.93 (0.85,32.64) | -3.40 (-9.99,3.25) | toneL:followingL | -193.70 (-234.18,-157.84) | -34.87 (-52.22,-18.31) | 11.35 (3.90,19.20) |
| conditionb | -82.37 (-95.85,-69.70) | - | - | toneL:genderM | - | - | -3.82 (-6.13,-1.42) |
| conditionc | -147.18 (-160.35,-134.27) | - | - | toneL:conditionb | 8.76 (-6.36,22.65) | - | - |
| genderM | -259.83 (-357.31,-175.79) | - | 10.24 (5.93,14.65) | toneL:conditionc | 22.90 (7.65,36.66) | - | - |
| voice+ | - | 23.49 (15.77,31.15) | -0.74 (-4.61,3.28) | conditionb:genderM | 14.59 (0.16,29.02) | - | - |
| stress | - | 19.64 (11.53,26.76) | - | conditionc:genderM | 77.30 (62.07,91.11) | - | - |
| vowela | 18.12 (1.53,35.29) | -19.09 (-27.74,-10.73) | -5.98 (-9.40,-2.72) | toneL:syll | 41.89 (24.61,58.76) | - | -7.52 (-11.16,-4.06) |
| vowele | 43.52 (25.27,63.39) | -0.07 (-7.63,6.84) | -6.48 (-10.40,-2.56) | previousH:voice+ | - | -3.37 (-12.29,5.77) | 6.54 (2.00,11.17) |
| voweli | 80.20 (61.25,99.32) | -22.97 (-30.72,-15.23) | -3.45 (-7.39,0.62) | previousL:voice+ | - | -34.44 (-43.15,-25.28) | 0.37 (-4.12,4.99) |
| vowelu | 64.85 (45.71,85.54) | -11.04 (-19.08,-3.22) | 3.17 (-0.99,7.18) | toneL:stress | - | -27.43 (-35.85,-19.53) | - |
| syll | -85.31 (-95.72,-75.43) | - | 2.97 (0.91,5.12) | previousH:toneL:followingL | 167.93 (111.06,225.88) | 14.56 (-6.53,39.84) | -21.35 (-33.37,-10.53) |
| | | | | previousL:toneL:followingL | 112.15 (63.54,160.37) | 42.82 (22.14,66.85) | -3.91 (-14.51,5.97) |

TABLE 2. Fixed effects and 95% Highest Posterior Density (HPD) confidence intervals for the three FPC score models. CovariateA1:CovariateB2 indicates an interaction between covariate A level 1 and covariate B level 2, assuming that the covariates have different categorial levels (otherwise the level is omitted). A "-" indicates that the effect not present in the model. l95 - lower 95% HPD interval boundary, u95 - upper 95% HPD interval boundary. It is important to note that the effects are on the scale of the eigenfunctions rather than relative to one another and therefore estimate value comparison across components is not meaningful.




(a) Mean Function and Eigenfunctions




(b) Estimated Covariance Function

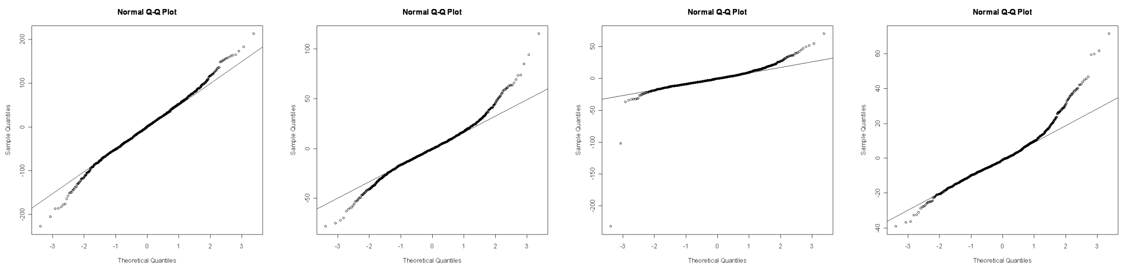
FIGURE 3. FPCA Analysis: Mean function and eigenfunctions for the first 3 components which account for 99.8% variance in the data, along with the estimated covariance function of the data. The estimated covariance function is smooth and thus additional presmoothing was not deemed necessary.

| Main Effect | FPC1<br>Estimate<br>(l95,u95) | FPC2<br>Estimate<br>(l95,u95) | FPC3<br>Estimate<br>(l95,u95) |
|---|---|---|---|
| speaker (sd) | 52.67<br>(31.79,109.39) | 7.19<br>(4.60,14.56) | 2.36<br>(1.27,5.04) |
| word (sd) | 31.13<br>(22.02,47.63) | 14.63<br>(11.08,22.49) | 7.96<br>(5.77,12.07) |
| residual (sd) | 55.93<br>(53.82,57.97) | 20.18<br>(19.60,21.12) | 10.85<br>(10.46,11.28) |

TABLE 3. Random effects (standard deviations) and 95% Highest Posterior Density (HPD) confidence intervals (of standard deviations) for the three FPC score models. See Table 2 for description of label meanings

97.0% of the variance in the data. In a limited analysis of the data, where only the median of each curve was modelled univariately with an LME (Evans, Chu, and Aston 2008), the model for the median coincided exactly with the model found for the FPC scores for the first component. On examination of the eigenfunction, this is not surprising. This eigenfunction is essentially flat, yielding a "shift" effect in the data, either up or down, depending on the covariates. However, it is reassuring to note, that despite allowing the contours to be non-parametrically specified, the first component did conform to expected linguistic theory for Luobuzhai Qiang in that the most important aspect of the tonal change is a shift rather than a contour change. In particular, the largest contributing covariates to the first eigenfunction were gender, tone, vowel type and sentence type. The random effects of subject and word item were both also significant. This indicates that the shift is speaker dependent, as well as dependent on the word item being said. While these effects are still relatively small in comparison to the effects of gender and tone, their significance shows that it is still important to consider the random nature of these effects in the analysis.

In many applications, with such a large percentage of the variance explained by one component, the modelling would cease here. However, in this data, as there are a large number of covariates, this would miss very important contour effects in the data, the primary purpose of the modelling. Indeed, it would be deemed that several important linguistic covariates did not affect F0. However, the second eigenfunction alters the start and end values of the contour without affecting to a great extent the middle of the contour. Many effects, such as the initial consonant, would only be expected to affect the beginning or end of the vowel. None of these "edge" effects were significant in the model for the FPC scores of component one (unsurprisingly given the flat nature of the contour). However, all the covariates which could be seen as edge effects are present in the model for the FPC scores of component two. In addition, some of the effects which were greatest in the first model, such as gender and tone, are either insignificant and thus excluded from the model or small in their own right but included in higher order interactions with edge effects causing them to remain present in the model. This shows the importance of considering a larger number of eigenfunctions when covariates are present. It was also of interest linguistically that in Qiang, it would appear that stress (here indicated by relative intensity) is an "edge" effect, rather than affecting the overall pitch level. This can be observed as it is only present in the model for the second eigenfunction.




(a) First FPC (b) Second FPC (c) Third FPC (d) Third FPC (no outliers)

FIGURE 4. FPCA Analysis: Diagnostic QQ plots for the different functional principal component score linear mixed effect models. The first looks acceptable, the second shows slight evidence of heavy tails although not strong, while the third shows strong evidence of outliers and skewness. After outlier correction, the plot is better, although evidence of skewness remains.

The third eigenfunction FPC scores have an associated LME model that is fairly similar (although not identical) to the LME model for the first eigenfunction FPC scores. However, the eigenfunctions themselves are very different in shape. This allows the contour to change in respect to these covariates in a way that is more complex than a pure pitch shift. Indeed, it is the previous and following tones (and interactions) that have the greatest magnitude coefficients in this third model, in contrast to the gender and tone effects in the first model.

The regression diagnostics (see Figure 4) looked fairly good for the first FPC score model, but became progressively worse for each of the subsequent components, which is unsurprising given that the amount of variation explained drops rapidly in each of the components, making them more susceptible to differing noise characteristics. However, given the Gaussian assumptions which underpin the decomposition, even though there was evidence of departures from Gaussianity, particularly in the third FPC score model, apart from the removal of obvious outliers (which were confirmed by outlier tests), corrections were not made. It would be of considerable interest to extend the model to account for some of these departures and this will be the subject of further research. Having said that, the Gaussianity assumption is fairly robust overall, as the first FPC score model contributes so much to the overall estimate of the curve.

A characterisation of the covariate effects on the F0 contour for Luobuzhai Qiang can be found by examining the overall model for the data. This model is made up of the non-parametrically defined mean function and the three eigenfunctions, and the parametric models associated with each of those functions. A prediction for any particular effect could be made by combining the output for all the models together. For example, the estimated curves in Figure 1 represent the male and female estimated curves for the word /ɕí tʂú ˈpiàn tsə̀/ ("riverbank"). The fit is close to the data, and is here plotted without the subject and word random effects being included, so as to see how an average word of the form of "riverbank" would be said. It is very noticeable that the form of each curve is highly dependent on the covariates. Indeed, this can yield additional insights into the linguistic structure of Luobuzhai Qiang. A high tone becomes elevated before a low tone, to the extent that it overrides the natural downtrend of the sentence (the second syllable is not lower than the first in the figure). Further, the male and female curves are not identically shaped (this is most noticeable in the two low toned syllables). While being male affects the first eigenfunction as expected, displacing F0 downward dramatically relative to the female curve (due to different pitch range for men and women), it also affects the second and third eigenfunctions, making a subtle difference in the shapes of the curves.

Overall, this entire functional response model provides a much richer yet still interpretable formulation for the natural utterances recorded than would be possible under a model based on a single point measure for each response.

## 6. Discussion

The statistical modelling and analysis of linguistic data is becoming ever more prevalent (Johnson 2008; Baayen 2008). However, typical methodology in phonetic analysis does not take into account the full quantitative effects of contour changes, either because the full contour is not modelled, or because a large number of restrictions are placed on the permitted utterances when the full contour is considered. This paper has presented a combined FPCA and LME model to account for the curve nature of the data, in the presence of a large number of possible covariates and interactions. The main advantage of this approach is the simplicity inherent in using the FPC scores to reduce the dimension of the functional responses. The covariates are presumed to affect the data through the FPC scores, and thus flexible yet understandable interpretation of the model is possible. While the use of scores as surrogate data has been previously suggested (Chiou, Müller, and Wang 2003), the complete nonparametric formulation used there limits the application of the model to covariates with dense structure, while also requiring the use of a single-index model, with its inherent problems of interpretation. The semiparametric approach undertaken here allows any covariate that can be modelled in an LME model to be modelled in this system too with the inherent advantages of relatively straightforward interpretation.

The data itself can be considered smoothed by the preprocessing step that was taken to determine the F0 curves. In part, the curves are smooth due to the quite rigourous experimental setup where the participants were trained to use the microphone, different from many speech processing applications. However, the curves are also smooth due to the intrinsic nature of the sound being produced, in that that in linguistic theory it is believed that due to physiological reasons, measurement interludes briefer than ten milliseconds are not likely to show meaningful changes in F0. Therefore, it is standard linguistic practice to use ten millisecond intervals or normalised data with intervals of approximately ten milliseconds. As normalised vowel time is used in this study, and the average vowel length was approximately 100 milliseconds, 10% intervals were taken. This certainly impacted on the smoothness of the data (as can be seen by the covariance function in figure 3) but it is unlikely that the data were over-smoothed for the reasons given above.

It might have been possible to use particular predefined bases for the functional data such as smoothing splines or wavelets. Indeed a polynomial basis would seem to be a good representation of the data given the eigenfunctions found (see figure 3). It would appear that the first three eigenfunctions would be well represented by a constant, linear and quadratic curve respectively. However, this was only possible to determine from the eigenfunctions post processing. There was no reason to apriori choose a polynomial base over any other, and thus the FPCA approach was preferred. In another language it is likely different bases would be required to model the data, and using the FPCA components at least guarantees the most parsimonious orthogonal representation. Given that the design of the experiment was fairly orthogonal itself, it is not then particularly surprising that the regression effects split between the different FPCA components, but it was interesting to see that in particular the first component represented "shift" and the second component represented "edge" effects.

One particular area that deserves further investigation is that of the relationship between the regression diagnostics and the model. While in principle the components are independent and consistently estimated, this is not true in finite sample data. The regression diagnostics for the third FPC in the example were not particularly good, and while it is likely that the approximation does not affect the end result to a great extent (given the small amount of variance of the signal explained by this component in any case), it would be more satisfactory to determine whether it is truly that the model does not fit, or whether the diagnostics need to be modified to account for the extra variation in the system.

It could also be argued that there is possible overfitting of the data as so many covariates were considered. Firstly, all covariates have been previously recognised as playing important roles in F0 production. Therefore excluding any apriori was not possible, and biasing towards a simpler model not necessarily a correct assumption, as many of the covariates were unrelated. Secondly, standard methodology in FPCA might well have deemed the "edge" covariates not present in the data, as so much variation was explained by the first FPC. However, these effects are associated with the second FPC scores, and as such some notice must be taken of the underlying linguistic theory in building the model, rather than taking a purely pragmatic statistical approach. It would have been of interest to reserve part of the data as a "test set" to investigate the predictive ability of the model, but given the very limited data available in typical phonetic fieldwork studies, including this one, it is not possible to do this and retain any particular confidence in the estimated model. However, it should also be understood that the primary purpose for the model was to try to determine a linguistic description of the language rather than to predict further utterances. It would have been of considerable interest to return to the Luobuzhai area to collect further data, using the model to design further experiments, but due to the Sichuan earthquake, this is now impossible.

The principle aim in the paper is the interpretability of the model, with particular reference to the linguistic data under analysis. This is slightly at odds with other speech recognition based procedures such as HMM methods (Rabiner 1989), where the primary aim is classification of the words themselves, rather than the analysis of the linguistic structure of the language. However, there is no reason that a successful characterisation of the language from the functional responses could not also be of use in speech recognition.

While we have concentrated on linguistic data analysis in this paper, the model presented could inherently be used in other applications where covariates could possibly affect curve data, but where non-parametric models of the covariates are not easily applicable. Indeed, while the methodology is likely to be fairly robust to departures from normality, by making use of similar models to the LME model, such as generalised linear mixed effect models, non-Gaussian data could be modelled in a very similar framework.

## Acknowledgements

The authors wish to acknowledge the great assistance of the following people; Chenglong Huang during the data collections; Mr. Ming-Jie Wang, the primary consultant in the decision of appropriate nouns; Man-ni Chu and Michael Jyh-Ying Peng for data preparation;

Prof. Chiu-yu Tseng and the Institute of Linguistics, Academia Sinica for equipment used in gathering the data. JE was supported by NSC(Taiwan), Grant No: 95-2411-H-001-077. JADA was supported by EPSRC and HEFCE through the CRiSM programme grant.

## Appendix A. Properties of Estimated FPCA scores

Assume the random process $Y(t)$ has the mean $\mu(t)$ and the eigenvalue-eigenfunction pairs $(\lambda_j, \phi_j(t))$ defined through the covariance operator. The Karhunen-Loève representation of the random process is $Y(t) = \mu(t) + \sum_{j=1}^{\infty} A_j\phi_j(t)$, where the FPC scores, $A_j = \int(Y(t) - \mu(t))\phi_j(t)dt$, are uncorrelated random variables with the mean of zero and the variance $\lambda_j$ satisfying $\sum_{j=1}^{\infty} \lambda_j < \infty$. When the random function $Y(t)$ follows a Gaussian process, it can be shown by the definition of $A_j$ that $A_j$'s are independent Gaussian random variables. Since $\mu$ and $\phi_j$'s are unknown, they are replaced with their estimates and the estimates of $A_j$'s are obtained by discrete approximations such that $\hat{A}_j = \sum_{l=1}^{m}(Y(t_l) - \hat{\mu}(t_l))\hat{\phi}_j(t_l)\Delta_l$, where $\Delta_l = (t_l - t_{l-1})$. Note that $\mu$ and $\phi_j$'s are consistently estimated with the uniform convergence rates provided in Yao, Müller, and Wang (2005) and the $L^2$ convergence rates in Hall, Müller, and Wang (2006), respectively, under certain regularity conditions on the design and number of time points, the number of curves and the relative order of bandwidths. Given the consistent estimates $\hat{\mu}$ and $\hat{\phi}_j$, it can be shown easily that $\hat{A}_j$ and $A_j$ are consistent. Further, under the Gaussian random process assumption, the estimated FPC scores, $\hat{A}_j$, for each $j$ follow the asymptotic Gaussian distribution.

## Appendix B. Words used in study

| No | Form | tones | Glossary |
|---|---|---|---|
| 1 | /dʐ̩ù ˈbè/ | LL | star |
| 2 | /dʐ̩ə́ ˈɕí/ | HH | day before yesterday |
| 3 | /ˈlí phò gè/ | HLL | trumpet |
| 4 | /tɕè ˈpjá ʁè gè/ | LHLL | corn cake |
| 5 | /pù qhà pà (gé)/ | LLLH | large intestine |
| 6 | /dʐ̩ò dʐ̩ò gé/ | LLH | ruler |
| 7 | /pú ʂ̩ú ʂtʂ̩à (gè)/ | HHLL | youth (n.) |
| 8 | /mù tɕhàn ˈthá mí/ | LLHH | robber |
| 9 | /ˈtshà tʂ̩ú qò qò/ | LHLL | storage room door |
| 10 | /ɕì ˈphú grə̀/ | LHL | root fibers |
| 11 | /ɕí tʂ̩ú ˈpiàn tsə̀/ | HHLL | river bank |
| 12 | /biá ɳ̩ú pì ˈqhuá/ | HHLH | female panda |
| 13 | /pù qhà ˈpà/ | LLL | large intestine |
| 14 | /ptú/ | H | flail |
| 15 | /tɕè ˈpjá/ | LH | corn cake |
| 16 | /tʂ̩hə̀ (ʂ̩)tə́ ˈquá/ | LHH | stomach |
| 17 | /biá ɳ̩ú ʐ̩dó/ | HHH | male panda |
| 18 | /pˤɹ/ | L | snow |
| 19 | /lì ˈχà sò gé/ | LLLH | tenderness |

Forms are given in International Phonetic Alphabet. No local writing system is available for Luobuzhai Qiang.

## References

Baayen, R. H. (2007). *LanguageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 0.4.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, special issue on Emerging Data Analysis Techniques*, in press.

Bates, D. M. and D. Sarkar (2007). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.9975-13.

Beckman, M. E. and J. Hirschberg (1994). The ToBI annotation conventions. Unpublished ms., Ohio State University and AT&T Bell Telephone Laboratories.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences Amsterdam 17*, 97–110.

Castro, P. E., W. H. Lawton, and E. A. Sylvestre (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics 28*, 329–337.

Chiou, J.-M., H.-G. Müller, and J.-L. Wang (2003). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society: Series B 65*, 405–423.

Crystal, D. (1990). *A Dictionary of Linguistics and Phonetics* (2nd ed.). Oxford, UK: Blackwell.

Di, C.-Z., C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi (2008). Multilevel functional principal component models. *Annals of Applied Statistics*, in press.

Evans, J. E., M.-N. Chu, and J. A. D. Aston (2008). Modeling of a two-tone system using linear mixed effects. *Journal of Phonetics*, in revision.

Evans, J. P. (2001). Contact-induced tonogenesis in Southern Qiang. *Language and Linguistics 2*, 63–110.

Faraway, J. (2006). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. New York: Taylor & Francis.

Ferraty, F. and P. Vieu (2006). *NonParametric Functional Data Analysis: Theory and Practice*. New York: Springer.

Fujisaki, H., W. Gu, and K. Hirose (2004). The command-response model for the generation of F0 contours of Cantonese utterances. In *7th International Conference on Signal Processing*, Volume 1, pp. 655–658.

Fujisaki, H. and K. Hirose (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Joumal of the Acoustical Society of Japan (E) S(4)*, 233–241.

Guo, W. (2002). Functional mixed effect models. *Biometrics 58*, 121–128.

Hall, P., H. G. Müller, and J. L. Wang (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics 34*, 1493–1517.

Johnson, K. (1999). *Acoustic and Auditory Phonetics*. Oxford, UK: Blackwell.

Johnson, K. (2008). *Quantitative Methods In Linguistics*. Oxford, UK: Blackwell.

Khouw, E. and V. Ciocca (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics 35*, 104–117.

LaPolla, R. J. and C. Huang (2003). *A Grammar of Qiang, with Annotated Texts and Glossary*. Mouton Grammar Library. Berlin: Mouton de Gruyter.

Liu, G. (1998). *Mawo Qiangyu Yanjiu [Research on Mawo Qiang]*. Chengdu: Sichuan Nationalities Press.

Mixdorff, H. (2000). A novel approach to the fully automatic extraction of Fujisaki model parameters. In *Proceedings of ICASSP 2000*, Volume 3, pp. 1281–1284.

Morris, J. S. and R. J. Carroll (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B 68*, 179–199.

Pinheiro, J. C. and D. M. Bates (2000). *Mixed-effect models in S and S-PLUS*. New York: Springer.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 257–286.

Ramsay, J. O. and B. W. Silverman (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Berlin: Springer.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (2nd ed.). Berlin: Springer.

Rice, J. A. and B. W. Silverman (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *Journal of the Royal Statistical Society: Series B 53*, 233–243.

Rose, P. (1987). Considerations on the normalisation of the fundamental frequency of linguistic tone. *Speech Communication 6*, 343–351.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. USA: Wiley.

Stanford, J. N. (2008). A sociotonetic analysis of Sui dialect contact. *Language Variation and Change*, in press.

Sun, H. (1981). *Qiangyu Jianzhi [A Brief Description of the Qiang Language]*. Beijing: Nationalities Press.

Trager, G. L. and B. Bloch (1941). The syllabic phonemes of English. *Language 17*, 223–246.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech 38*, 253–265.

Wen, Y. and M. Fu (1943). Wenchuan Lobuzhai Qiangyu yinxi [Phonology of the Qiang Language, Group II, Lo-pu-chai dialect]. *Studia Serica 3*, 14–25.

Xu, C. and Y. Xu (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association 33*, 165–181.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics 27*, 55–105.

Xu, Y. (2006). Principles of tone research. In *Proceedings of International Symposium on tonal aspects of languages 2006*, La Rochelle, France, pp. 3–13.

Yao, F., H. G. Müller, and J. L. Wang (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics 33*, 2873–2903.